Contents

1. Introduction	1
2. Literature Review	3
3. Rationale	7
4. Objectives	7
5. Research Methodology	8
6. Expected Outcomes	9
References	10

1. Introduction

In new era of online published news articles hundreds of articles published every day by the popular news agencies like CNN, BBC etc. The Amount of online published news articles increases day by day, so the reader need to find news article related to specific event type of his or her interest. In general reader to find news articles of his or her interest, a reader filter out desired into related to event of interest from headlines and teaser by scanning various sources of news articles about the same news information he already know about it. Therefore people always want to extract comprehend a huge amount information in very less time. A reader not willing to read through a collection of news articles needs a representation of the news in a compact from describing the event type briefly.

The amount of unstructured data enormously increases every day, so does require a process it automatically extract different types of knowledge from data.[1,2] A continuously occurred problem in processing this large amount of data in intelligent manner because most of them data are unstructured form e.g. data written in human understandable language. These information in the form of online broadcast news, blogs to comments and social communication data. [16, 17, 22] Thus, it may also consist of various types of unrelated data or noisy discussions. Event extraction is the task to identifying the key terms related to the specific events from unstructured data.

Event is a specific occurrence involving participants. An Event is something that described as a change of state, so in various fields are classified in terms of events. Formally, the task of event extraction is to automatically identify events in collection of text and to derive detail information about them is known as event extraction. Event extraction from texts could be beneficial to various domains. In the online published news articles application, the ability to identifying events may enhance the performance of specific news recommendation. Since the online published news articles may be selected more efficiently based on the specific extracted events and user preferences. [11, 12, 13] Event Extraction and

Categorization of online published news articles have many application such as government policy makers uses the information related to significant international incidents tracking to make better decisions regarding policies. In cyber security domain, ongoing cyber-attack incidents help to monitoring. Another most valuable application of event extraction in the area of stock market extremely sensitive to the breaking news for deciding the stock trading plaining such as time, price and volume.

Formally, events are extracted from text at sentence level, some past studies are showed that the extraction of events using relations, entity and document level. Event Extraction are crucial and quite challenging task, as the argument detection and classification heavily depends on it. [21, 25, 27] in the news articles can be challenging to identified as the same event might have in different from, either present in single and multi-word units.

Event extraction is comes from the unstructured data where the large number of non-event examples are present in the other so making event and argument identification are more challenging task.

For such essential characteristics of an online published news articles document required more semantic information processing, so that require to build a larger scale news semantic knowledge base. After that such knowledge base apply the event extraction and event categorization is more challenging task.

2. Literature Review

In the field of event extraction and analysis many researcher have made extensive efforts to develop the several methodologies of event extraction from news articles. A numbers of efforts have been undertaken to address the analysis of narrative work in online news articles.

Chen et al [1] propose a dynamic multi pooling conventional neural network model (DMCNN). Which uses dynamic multi pooling according to event trigger and argument to receive more information. Michal Tillo et al [2] present a web toolkit to show the automatically computed suggestion are used to guide exploration process and also demonstrate how user extractors with one relation is identified.

Ralph Grishman et al [3] describe a two-stage training algorithm for neural networks that effectively transfers knowledge from the other event types to the target type. They also investigate the effectiveness of the method compare to other state of art method. This method for extracting of knowledge may be more efficient use of real time dataset for description.

Yagung et al [4] present a timeline that shows significant events about query word based on our corpus, which can inform user's important things they are interested in. In the procedure to cluster sentences by their dates, they considered the value of no accurate date sentences. Because it is now one date vector for one date, they can only extract the most important thing in each date vector. In reality, it is possible that a number of things with same type happened on same date.

Shulin liu et al [5] propose to exploit argument information explicitly for ED (Event Detection) via supervised attention mechanisms. In this proposed model they systematically investigate the proposed model under the supervision of different attention strategies.

Somayeh Keshavarz et al [6] proposed a novel probabilistic inference framework for complex video event recognition using supervised action concepts. In these proposal to attempt to model the conditional relationships between complex events and the exhaustive set of intermediate concepts by constraining dependencies to pairwise joint distributions while avoiding the need to manually re encode new graph structures as the number of concepts increases.

Mahdi Namazifar [7] suggest a Named Entity Recognition (NER) method aims at locating and classifying named entities in text. Proposed NER method apply to find Named entity from Tweets. Researcher suggest that the NER is also used for the online news articles text.

Gills jacoab et al [8] present a dataset and classification experiments for company specific economic event Detection in English news articles. There is still plenty of room for improvement: more annotated data and augmentative resources are needed to further offset ambiguous event expressions.

Felix Hamborg et al [11, 12, 13] proposed open source system that retrieve answers in the questions form of what, when, who, where, why and how to describe online published news articles information in main events. In these researcher identified only the information related to main event they does not categorize all types of main events.

Quan Yang et al [14] present to obtain event profiles event extraction automatic framework using upsuperviesed learning method. Their proposed framework based on event profiling. For implementation of work lots of challenges occur such as find out event types in given event initially unknown for users, lake of event types schemas in previously defined in model. So that for event profiling to their event types is still demanding in field of information extraction from text.

Haibo Ding et al [16, 17, 22] present a method based on word and contexts for Human Needs categorization. They develop a method for assign words to related concept in between of categorization process. Researchers are create supervised model for labelling the words into their sematic concepts for human need information categorization. They use 10 types of semantic concepts in their work.

J. Walkar et al [19] present a new model which is based on GRU that combines the temporal structure information and syntactic information in that mechanism. They result show that it is competitive with other state of arts methods like neural network architectures using empirical evaluations under the split of ACE dataset. They suggest that event can be categorize using their types.

Shuo Yang et al [21] propose an efficient collapsed Gibbs sampling approach to infer the truths of news and the users' credibility without any labelled data. Experiment results on two datasets show that the proposed method significantly outperforms the compared unsupervised methods.

Rui Wang et al [25] present a novel approach based on adversarial training to extract the structured representation of events from online text. The experimental comparison with the state-of-the-art methods shows that AEM achieves improved extraction performance, especially on long text corpora with an improvement of 15% observed in F-measure. In future work, we will explore incorporating external knowledge (e.g. word relatedness contained in word embedding) into the learning framework for event extraction.

Momna Nassem et al [27] proposed approach in which uses two level of clustering. In the first level of clustering identifies major events among diverse social media text, and the second level of clustering method sub events identifies to a given super event by using semantic relationship and other temporal information. After that they compare their results in terms of accuracy compare to other state of arts methods.

Above related work shows the study different methods and data set issue of previously used method for event extraction and categorization. There are some research gaps in this area, as:

- They assume that all words in a document are generated from a single event However long texts such news articles often describe multiple events which clearly violates this assumption.
- The domain specific data set are required for effective event extraction.
- Salient trigger and argument detection form article is still a challenging task.
- The semantic relationship is missing in trigger and argument due to that relevant event type are uncertain in nature.
- In Present event extraction and categorization task are biased in nature due to categorization of event some specific type only for analysis of special purpose only so the models are imperfect to all users.
- More precise event extraction and categorization of event type are required.

3. Rationale

Event extraction and Categorization from online news articles is distinct from other traditional event extraction task such as trigger and argument detection in text, classify the trigger and argument present in text etc. however these tasks do not provide categorization of event type from the events. The current state-of-arts methods [21,22,23,25,27] are developed on the assumption all words in a document are generated from a single event besides long texts such news articles often describe multiple events which clearly violates this assumption. However before extracting the events first check the given event are relevant to the argument and trigger or not and after that categorize these events to specific event type. To the best of my knowledge, we did not find any work which addressed this issue. Also, current technique largely focused on the event extraction tasks. However less effort is made on the understanding the categorization of event types.

The existing research in event extraction and categorization not much focus on the categorization of events present in the text. These opportunities motivate us to propose our research work on the event extraction and categorization.

4. Objectives

The main objective of our research proposal is to development of efficient supervised classification method for automatic categorization of online published news articles, for achieve main objective have following sub objectives are:

1. Develop a method to identify event sentences from online published news articles by using triggers and arguments present in sentences.
2. To develop a sequence labelling machine learning model for extraction of Named Entity preset in event sentences.
3. To develop a method for identification of contextually similar event sentences.
4. To develop an improved supervised classification method which automatically categorizes article based on identified event with respect to event type.

5. Methodology

The proposed methodology for the proposed work as follows:

- Data Collection – In the data collection phase, we will derive the datasets from existing knowledge bases such as Wikipedia, YAGO and DBpedia. These knowledge bases contains collection of important public events occur in the past. These events can be used for learning the identification of event from the text.
- To extract events from text, we first need to learn the extraction from the past events. For learning, we will use existing machine learning libraries such as Microsoft azure.
- To extract the context of an event, we will extract named entities from the event sentences. We will develop sequence labelling machine learning method for this task. We will use entity labeled data such as ACE, Stanford named entity datasets etc for the training of the machine learning method.
- To determine the similarity in between events, we will develop method which uses the event context (named entities, triggers, arguments).
- Event Categorization – We will develop supervised classifiers by learning the event types of existing events in the datasets. The learned classifier will be used for classification of unlabeled events.

6. Expected Outcomes :

- Provide huge information to specific domain at one place like crime and law, sports etc.
- Its help to construction to knowledge graph for system like question answering system.
- More efficient information available for fact checking, opinion mining and future trade analysis.
- Provide efficient information to find news article to his or her interest.

References

1. Yubo Chen, Liheng Xu, Kang liu, Daojian Zeng, and Jun Zhao. “Event Extraction via dynamic multi-polling convolutional neural networks.” *In ACL-IJCNLP2. 2015 pp 167-176.*

2. Thilo Michael and Alan Akbik. “A web toolkit for exploratory relation Extraction.” *In ACL-IJCNLB, 2015.*

3. ThienHuu Nguyen, Lisheng Fu, Kyunghyun Cho and Ralph Grishma. “A two-stage approach for extending event detection to new types via neural network.” *In WRepL4NLP, 2016 pp 158-165.*

4. Yaguang Wu, Haichun Sun and Chuhuang yan. “An Event Timeline Extraction Method based on News Corpus.” *International Conference on Big Data Analytics 2017.*
5. Liu, Shulin, et al. "Exploiting argument information to improve event detection via supervised attention mechanisms." *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics 1 2017.*

6. Keshaverz Somayeh, Saleem Imran, Atia George. “Exploiting Probabilistic Relationships between Action Concepts for Complex Event Classification.” *IEEE Transaction and Information Theory, 2017.*

7. Mahdi Namazifar. “Named Entity Sequence Classification.” *arXiv preprint arXiv. 1712 (2017).*

8. Gills Jacobs, Els Lefever and hoste. “Economic Event Detection in Company Specific News Text.” *In the proceedings of first workshop on Economic and NLP, Melbourne Australia 2018 ACL page 1-10.*

9. Mohammad Kamel, Neda keyvani, “Sentiment Content Analysis and Knowledge Extraction from News Articles.” *Association of computer linguistics conference 2018 page 13-18.*

10. Sonit Singh. “Natural Language Processing for Information Extraction.” *Association of computer linguistics conference 2018 page 13-18*

11. Felix Hamborg, Corrina Breitinger, Mortiza Schboutz and Bela Gipps. " Extraction of Main Event Descriptor from News Articles by Answering the journalist Five W and One H Questions." *FCDL'18 June 3-7 2018, Fort Worth, TX, USA.*

12. Felix Hamborg, Corrina Breitinger, Mortiza Schboutz and Bela Gipps. "Giveme5W: Main Event Retrieval from News Articles by Answering the journalist Five W." *iConference 2018 in Sheffield , UK.*

13. Felix Hamborg and Bela Gipps. "Automated Identification of media bias in news articles: an interdisciplinary literature review." *International Journal on Digital Library, Springer 16 November 2018.*

14. Quan Yang, Xiang Ren Wenqi He, Chao Zhang and heng ji et al. "Open Schema Event Profiling for Massive News Corpus." *CIKM'18, October 22-26 2018,Torina, Italy,*

15. Navya Yarrabelly and Kamalakar Karlapalem. "Extracting Predictive Statements with their Scope from News Articles." *Proceedings of 12th International AAAI Conference on Web and Social Media (ISWSM 2018).*

16. Haibo Ding, Tianyu Jiang, and Ellen Riloff. "Why is an even affective? Classifying affective events based on human needs." *In the AAAI-18 Workshop on Affective Content Analysis, 2018.*

17. Haibo Ding and Ellen Riloff. "Human needs categorization of Affective events using labeled and unlabelled data." *InProceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human LanguageTechnologies (NAACL 2018).*

18. P. Sarma, Y. Liang, and W. Sethares. "Domain adapted word embedding's for improved sentiment classification." *In Proceedings of the ACL 2018 Workshop on Deep Learning Approaches for Low-Resource NLP.*

19. J Walker Orr, Prasad Tadepalli and Xiaoli Fern. "Event Detection with Neural Network: A Rigorous Empirical Evaluation." *arXiv preprint arXiv. 1808 (2018).*

20. P.K. Choubey, K. Raju, R. Huang. ", Identifying the most dominant event in a news article by mining event coreference relations." Proceedings of the 2018 Conference Of the North American Chapter Of the Association For Computational Linguistics: Human Language Technologies, vol. 2, 2018, pp. 340e345. Short Papers.

21. Yang S, Shu K, Wang S, Gu R, and Liu H. "Unsupervised fake news detection on social media: A generative approach." *Proceedings of 33rd AAAI Conferences on Artificial Intelligence, 2019.*

22. Ding Haibo, Riloff Ellen and Feng Zhe. "Improving Human Needs Categorization of Events with Semantic Classification." *Proceeding of Eighth Joint Conference on Lexical and Computation Semantics by ACLJune 2019.*

23. Y Belinkov, J Glass. "Analysis methods in natural language processing: A Survey." *Transaction of the Association for Computational Linguistics, 2019, vol 7, pp 49-72.*

24. Anais Ollagnier and Hywel Williams. "Classification and Event Identification Using word Embedding." *Transaction of the Association for Computational Linguistics, 2019, vol 7, pp 91-122.*

25. Rui Wang, Deyu Zhou and Yulan He. “Open Event Extraction from Online Text using a Generative Advisable Network.” *Proceedings of 33rd AAAI Conferences on Artificial Intelligence, 2019*

.

26. Fetemeh Torabi and Maite Toboada. “Big Data and Quality Data for fake news and misinformation detection.” Big Data and Society, January – June 2019. Pp 1-4.

27. Momna Nassem, Basit Shafiq Soon Chon, Shafay Shamail. “Discovering Events from Social Media for Emergency Planning.” Association for Computing Machinery, June 18, 2019, Dubai, UAE.

28. Suchanek FM, Kasneci G, Weikum G, "Yago: a core of semantic knowledge” *in Proceedings of the 16th international conference on World Wide Web ACM, 2007. 697-706.*

29. Sören Auer, Christian Bizer, Georgi Kobilarov, Jens Lehmann, Richard Cyganiak, and Zachary Ives, "Dbpedia: A nucleus for a web of open data” *in the semantic web. Springer, 2007. 722–735.*